Phylogenetic analyses of Teleki grapevine rootstocks using three chloroplast DNA markers

Poczai, Péter

2013

Poczai , P , Hyvönen , J , Taller , J , Jahnke , G & Kocsis , L 2013 , ' Phylogenetic analyses of Teleki grapevine rootstocks using three chloroplast DNA markers ' , Plant Molecular Biology Reporter , vol. 31 , no. 2 , pp. 371-386 . https://doi.org/10.1007/s11105-012-0512-9

http://hdl.handle.net/10138/324957
https://doi.org/10.1007/s11105-012-0512-9

Plant Mol Biol Rep (2013) 31:371–386
DOI 10.1007/s11105-012-0512-9

ORIGINAL PAPER

Phylogenetic Analyses of Teleki Grapevine Rootstocks Using Three Chloroplast DNA Markers

Péter Poczai · Jaakko Hyvönen · János Taller · Gizella Jahnke · László Kocsis

Published online: 12 September 2012

Abstract Teleki rootstocks are used in grapevine-producing countries all over the world. They represent one of the largest groups of available rootstocks but their origin is still in dispute although they have been regarded as *Vitis berlandieri* × *V. riparia* hybrids. To investigate their possible origin, we amplified and sequenced three chloroplast regions, two non-coding spacers (*trn*L-F, *trn*S-G) and the *trn*L group I intron in a core collection of Teleki rootstocks representing widespread accessions and related wild North American grape species (*V. berlandieri*, *V. riparia* and *V. rupestris*). Concatenated sequence data coupled with microstructural changes discovered in the chloroplast regions provided data to trace the maternal ancestry of the Teleki lines. All chloroplast regions showed both nucleotide and length variation. Length mutations in the non-coding regions represented mostly simple sequence repeats of poly-A and -T stretches. These indel characters exhibited additional diversity comparable with the nucleotide diversity and increased resolution of the phylogenetic trees. We found that a group of Teleki accessions position together with the wild grape species *V. riparia*. Another group of Teleki rootstocks formed a sister group to the other North American species *V. berlandieri*. These clades had moderate support values, and they do not share ancestry with other accessions of Teleki rootstocks resolved with high support value in the *V. riparia* clade. It seems that Teleki-Kober 5BB and 125 AA accessions might have a *V. berlandieri* maternal background. We also found great differences within putative clones of Teleki 5C and Teleki-Kober 5BB suggesting that the selection of these accessions was performed on heterogenous or mislabeled plant material collectively maintained under these names.

Keywords Grapevine · Group I intron · Non-coding spacer · North American wild *Vitis* species · *trn*L-*trn*F · *trn*S-*trn*G

Electronic supplementary material The online version of this article (doi:10.1007/s11105-012-0512-9) contains supplementary material, which is available to authorized users.

P. Poczai (✉) · J. Hyvönen
Plant Biology, Department of Biosciences, University of Helsinki,
PO Box 65, 00014 Helsinki, Finland
e-mail: peter.poczai@gmail.com

J. Taller
Department of Plant Science and Biotechnology,
Georgikon Faculty, University of Pannonia,
Festetics 7,
8360 Keszthely, Hungary

G. Jahnke
Research Institute for Viticulture and Oenology,
Centre of Agricultural Sciences, University of Pannonia,
Római út 181,
8261 Badacsonytomaj, Hungary

L. Kocsis
Department of Horticulture, Georgikon Faculty,
University of Pannonia,
Festetics 7,
8360 Keszthely, Hungary

Introduction

European vineyards were destroyed to large extent by grape phylloxera (*Daktulosphaira vitifoliae* Fitch, 1855) and became dependent on North American *Vitis* species (Snyder 1937; Granett et al. 2001). Large number of phylloxera-resistant North American *Vitis* rootstocks of diverse species were shipped to France in the 1870s (McLeRoy and Renfro 2008). Many interspecific hybrids resulted in large number of rootstocks that were grouped by ampelographic techniques (Galet 1967; Pongrácz 1988). It is very difficult to distinguish these accessions phenotypically (Fig. 1) because of their common parentage from *Vitis riparia* Michx., *V. rupestris* Scheele and *V. berlandieri* Planch. [Syn.: *V. cinerea* var. *helleri* (L.H.Bailey) M.O.Moore] (Walker and Boursiquot 1992). At first isozymes were used to

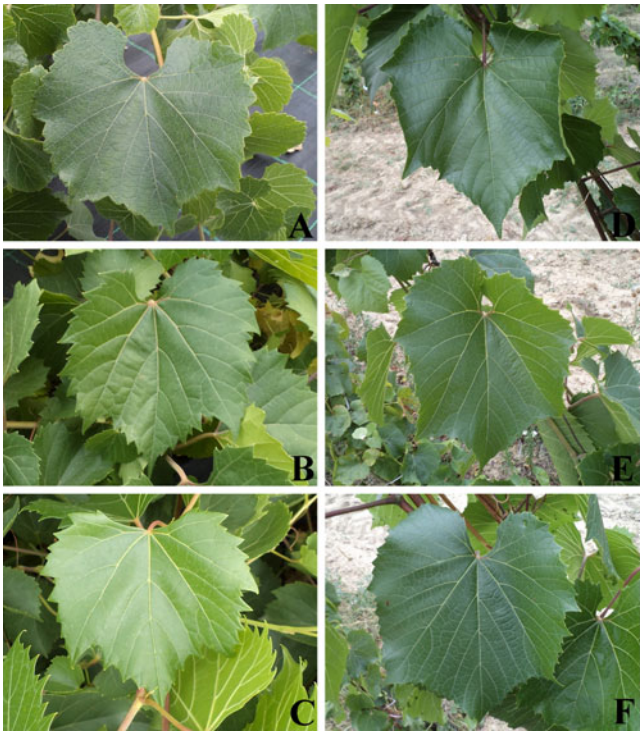

Fig. 1a–f Representative pictures of leaves of North American wild *Vitis* species and some Teleki rootstocks. **a** *V. berlandieri*: leaf is similar to that of *V. riparia*, except that teeth are much smaller and the tomentum is much denser at the tip of the shoots. **b** *V. riparia*: leaves are glabrous, three-shouldered to three-lobed cordiform blades with cordate bases. **c** *V. rupestris*: leaves are small glabrous, often three-shouldered, rarely shallowly three-folded; blades are reniform and folded invards; base truncate with acute to short acuminate apices. **d** Teleki–Kober 5BB: the petiolar sinus is prominently open with a characteristic upturned and wavy leaf margin. **e** Teleki–Kobber 125AA: leaves are gross, unlobed, with dark green colour with a widely opened sinus. **f** Teleki 5C—the rootstock used most widely: leaves are mostly unlobed, dark green; leaf rims curved upwards and sharply dented

distinguish grape rootstocks (Walker and Liu 1995), but with fairly poor success. Later, RAPD (randomly amplified polymorphic DNA) markers proved to be more successful (Xu et al. 1995), and RFLP analyses have also been performed on 16 *Vitis* accessions and clones (Bourquin et al. 1991). Hong and Walker (1998) published the results of efficient use of microsatellites for identification. Since then, SSRs have been exploited extensively in a number of countries for identification of cultivars, characterisation of grape genetic resources (Fatahi et al. 2003; Hvarleva et al. 2004; Grando and Frisinghelli 1998; Moravcová et al. 2006), verification of synonyms or homonyms (Fossati et al. 2001; Labra et al. 2001), parentage analysis (Bowers et al. 1999a; Sefc et al. 1998a) or genomic mapping (Adam-Blondon et al. 2004; Fisher et al. 2004; Riaz et al. 2004). Traditional cultivars of *V. vinifera* L. were genotyped by SSR markers but little data is currently available for rootstocks (Sefc et al. 1998b; Fatahi et al. 2003; Dzhambazova et al. 2007; Upadhyay et al. 2007; Cseh et al. 2006).

One of the largest groups among the rootstocks are the so-called Teleki lines. Due to the poor performance of the available phylloxera-resistant rootstocks on soils with high lime content in 1890s, the Hungarian farmer Zsigmond Teleki decided to develop his own lines. He purchased *Vitis berlandieri* canes from France. However, their import was prohibited because of "black rot" disease. Therefore Teleki ordered 10 kg grape seeds from southern France, from Mr. Rességuier (Teleki 1927). According to Teleki (1927), those seeds derived from open pollination of *V. berlandieri*. From 40,000 obtained seedlings showing large morphological variation corresponding to different species of *Vitis*, Teleki selected all seedlings resistant to fungal diseases, tolerant to high lime content, and with the main morphological characters of *V. berlandieri* × *V. riparia*. He was able to classify the plants into ten groups based on morphological traits. Groups 1, 2 and 3 included all the *V. vinifera* type plants, groups 3, 4 and 5 were more like *V. riparia*, groups 6, 7, and 8 were more like *V. berlandieri* and group 10 was reminiscent of *V. rupestris* (Bakonyi and Kocsis 2004). Within the groups, he marked the plants with the letters 'A' and 'B' according to the colour of the shoot apex and sex of the flowers. In this way, he chose what he considered to be the most useful groups, e.g. 4A, 5A and 8B. Two of these groups, 8B and 5A, were marketed for quick propagation, but these did not represent progeny of individual plants. Later, he selected T5C and T8B as single plants with defined traits from these pools. Further selection from Teleki's material was necessary and was performed by other breeders, such as Fuhr and Rodrian for SO4, and Kober for K5BB and K125AA (Bakonyi and Kocsis 2004; Manty 2005). The Teleki rootstocks are now used widely in viticulture in almost every wine-growing region and therefore it is very important to know if sensitivity against phylloxera resistance is present or not in the rootstocks currently used.

Morphological and growth characteristics of Teleki rootstocks are well documented, but their genetic background is still unknown. The clarification of the origin of Teleki lines has not only scientific but practical value as well, because the growth features of the rootstocks and on the grafted scions are under strong genetic control. The effectiveness of breeding correlates with the genetic background of the samples. This is true especially for woody plants such as grapevine, where 3–5 years are needed to uncover the results of breeding. This is why the clarification of the origin and relationships of these accessions is highly important. The aim

of this study was to improve our knowledge about the maternal genetic background of Teleki lines using non-coding chloroplast sequences.

Materials and Methods

Plant Material, PCR Amplification and Sequencing

DNA was extracted from cryopreserved leaves using the CTAB method of Doyle and Doyle (1987), incorporating the modifications by Poczai et al. (2009). Details of the plant material used in this study are provided in Supplementary Table 1. Three chloroplast regions were PCR amplified in two steps. The *trn*L-*trn*F region, containing the *trn*L^{UAA} gene, its intron and the 3′*trn*L^{UAA}-*trn*F^{GAA} spacer was amplified in one overlapping fragment with the universal primers *C* (5′-CGA AAT CGG TAG ACG CTA CG-3′) and *F* (5′-ATT TGA ACT GGT GAC ACG AG-3′) designed by Taberlet et al. (1991). These primers were used for sequencing coupled with primers *D* (5′-GGG GAT AGA GGG ACT TGA AC-3′) and *E* (5′-GGT TCA AGT CCC TCT ATC CC-3′) also by Taberlet et al. (1991) in cases when reads were too short to generate reliable sequences. The intergenic spacer spanning through the *trn*S^{GSU} and *trn*G^{UCC} transfer RNA genes were amplified with primers trnS (5′-GCC GCT TTA GTC CAC TCA GC-3′) and trnG (5′-GAA CGA ATC ACA CTT TTA CCA C-3′) (Hamilton 1999). These primers were used also for direct sequencing. Amplification reactions were performed in 10 µl volumes containing: 5 µl NFW (nuclease-free water), approximately 20 ng template DNA, 50 µg bovine serum albumin (BSA), 0.5 µM of each primer, 0.2 mM dNTP, 5 µl 10×DreamTaq buffer and 0.5 U DreamTaq DNA Polymerase (Fermentas, Vilnius, Lithuania). All reactions were performed in a MasterCycler ep384 (Eppendorf, Germany) with the following settings: 2 min at 94 °C for initial denaturation, 35 cycles of 30 s denaturation at 94 °C, 1 min annealing at 50 °C for *trn*L-*trn*F and 52 °C for *trn*S-*trn*G, and 2 min extension at 72 °C, followed by a final extension for 5 min at 72 °C. Amplification products were separated on 1.5 % agarose gels in 0.5 × TBE buffer (220 V, 0.5 h) and stained with ethidium bromide. Cleanup of PCR products was performed by removing non-incorporated primers with 10 U exonuclease I and degradation of nucleotides by 1 U thermosensitive alkaline phosphatase (Exo I and FastAP, Fermantas). PCR mixes were incubated at 37 °C for 15 min and the reaction was stopped by heating the mixture at 85 °C for 15 min. Direct sequencing was performed in an ABI 3130XL automated sequencer in both directions using the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit v.3.0. All sequences were annotated and deposited in NCBI GenBank (http://www.ncbi.nlm.nih.gov/genbank/). Accession numbers are available in the Supplementary Table 1.

Sequence Assembly, Alignment and Indel Coding

Forward and reverse sequence reads for all individuals were assembled into contigs using CodonCode Aligner v.3.7.1. (http://www.codoncode.org) generating a single consensus sequence for each terminal. Discrepancies were resolved by manually editing the traces using the compare option of the advanced assembly function. Final contigs were extracted from the compared assemblies and consensus sequences were converted to FASTA format. Multiple sequence assemblages were aligned with MUSCLE (Edgar 2004) as implemented in Geneious v.4.8.5 (http://www.geneious.com/) using default settings. Exon-intron boundaries and non-coding regions were identified and annotated in the alignments using the *Vitis vinifera* (NC007957) chloroplast genome as a reference. This facilitated additional editing and processing of the sequences. Further alignments were made by removing ambiguous positions and gaps with the command line version of Gblocks (Castresana 2000). Highly stringent settings were applied where all flanking sequences were conserved, and smaller (< 20 bp) blocks were discarded. Final datasets were concatenated and formatted to different extension files (FASTA, NEXUS) using the export options of Geneious. Sequence statistics were calculated either with SeqState (Müller 2005) or MEGA v5 (Tamura et al. 2011).

Gaps (or "indels") have been recognized widely as a valuable source of data for phylogenetic inference (Simmons et al. 2007). In order to utilize this source we coded indels based on microstructural changes. In the case of the concatenated alignments indels were incorporated as binary data according to the simple indel coding (SIC) method of Simmons and Ochoterena (2000) implemented in SeqState (Müller 2005). The program generated a ready-to-use NEXUS-file containing the binaries which were edited for further analyses.

Phylogenetic Analyses

Parsimony Analyses

We performed phylogenetic analyses with parsimony as an optimality criterion using the program Nona (Goloboff 1994) within a winclada (Nixon 2002) shell. The concatenated matrix of MUSCLE alignment + binary indels was reduced to 87 terminals by deleting all duplicate terminals. We performed five separate analyses (using processor time as a seed to randomise the order of the terminals) with the following settings: hold 3,000 (holding defined number of trees), 100 replications (search performed with multiple

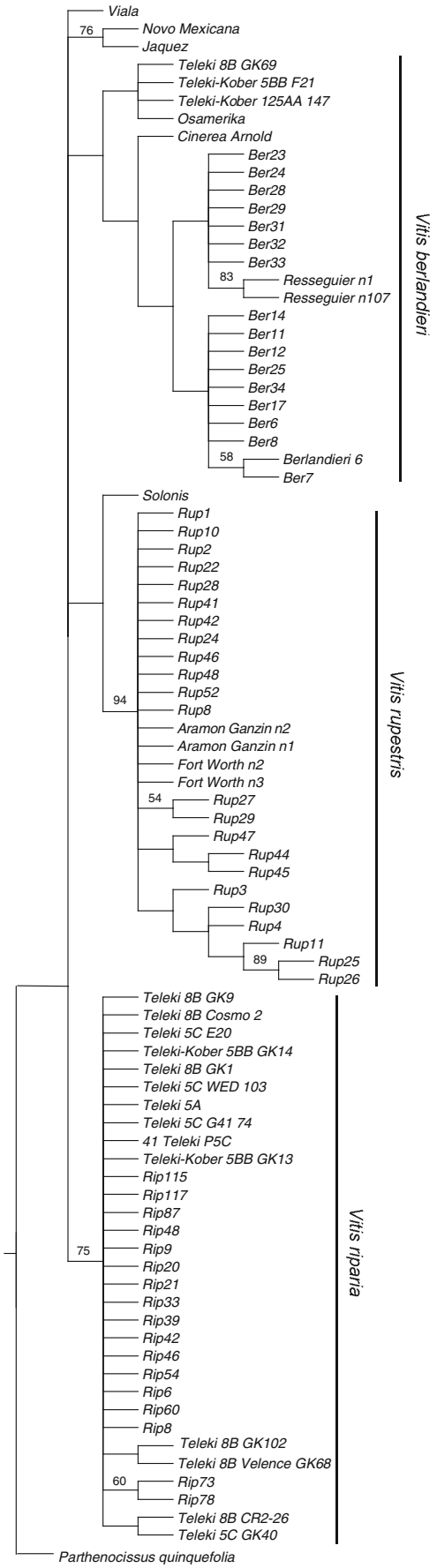

Fig. 2 The resulting strict consensus tree of the parsimony analysis. *Numbers* above branches represent jackknife support values

tree-bisection-reconnection algorithm mult*max*), hold/20 (keeping 20 starting trees for each replication). In addition, we performed also a larger analysis by holding up to 30,000 trees (hold 30,000) but keeping only 2 starting trees for each replication (hold/2). Jackknife (Farris et al. 1996) support values were calculated using 1,000 replications, with 10 search replications (multi*10) and with 1 starting tree per replication (hold/1).

Bayesian Analyses

Prior to Bayesian inference (BI) analysis, we identified best-fitting substitution models for the concatenated and separate datasets using the Akaike information criterion (AIC; Akaike 1974) as implemented in MEGA v5. Bayesian analysis was performed with MrBayes v3.2 (Huelsenbeck and Ronquist 2001), applying the Hasegawa-Kishino-Yano model (Hasegawa et al. 1985) with a proportion of invariable sites and gamma-shaped distribution of rate variation across sites (HKY+I+Γ) and the restriction site model for the binary indel partition. Posterior probability (PP) distributions were created using the Metropolis-coupled Markov chain Monte Carlo $(MC)^3$ method with the following assembled datasets: (1) concatenated MUSCLE alignment, (2) concatenated MUSCLE alignment + binary indels, (3) concatenated Gblocks alignment, and (4) concatenated Gblocks alignment + binary indels. We attempted to sample all trees that had a reasonable probability given the assembled datasets. For this reason we kept all sequences in the alignments even if they had identical sequences. This was based on population genetic reasons not to mislead the BI assuming a larger population. An initial analysis employing only two separate runs and four chains with 10^6 generations was conducted to fine-tune parameters. This preliminary analysis indicated that the default parameter for heating (*temp*= 0.2) results in infrequent swapping of states between the heated and cold chains. We found that *temp*=0.1 results in adequate states swaps, which we used for longer chains. Further analyses were initiated with four runs and eight chains (10^6 generations each) sampling every 100th generation. Runs were conducted on the computers of CSC-IT Center for Science (http://www.csc.fi), Espoo, Finland by assigning each chain to a separate CPU. Simulations were run until it was necessary to reach stationarity assessed from the average standard deviation of split frequencies <0.01. MC^3 convergence was explored by examining the Potential Scale Reduction Factor (PSRF) for all parameters in the model and plots of log-likelihoods over time together with other plots for all parameters allowed by Tracer v1.4 (Rambaut and Drummond 2007). Additional tests of convergence were conducted with the online program AWTY

(Nylander et al. 2008) using the 'cumulative' and 'compare' functions. The evidence of convergence can be understood as no significant changes in posterior probabilities over millions of generation, although true convergence can never be proven (Nylander et al. 2008). The states of the chains sampled before stationarity (split freq. >0.01) were discarded as burn-in (25 %). Trees from BI analyses were summarized as majority-rule consensus trees and edited with TreeGraph2 (Stöver and Müller 2010).

Maximum Likelihood Analyses

Phylogeny was also inferred using the maximum likelihood (ML) approach implemented in RAxML 7.2.6. (Stamatakis 2006). All runs were performed with the graphical user interface raxmlGUI 0.93 (Silvestro and Michalak 2012). The reduced MUSCLE alignment without binary coded indels was used for the analysis, since concatenated matrices are not supported by the software. Thorough bootstrap searches (1,000 replicates) were performed under the default general time reversible model of nucleotide substitution with rate heterogeneity following a discrete gamma distribution (GTR+ Γ). RAxML implements only the GTR model and is therefore applied in our analysis. Throughout this paper, 70–84 % bootstrap support is considered moderate and 85–100 % as strong support.

Sequence and Indel Variability

The concatenated MUSCLE alignment was analyzed with DnaSP v5 (Librado and Rozas 2009) to estimate polymorphism. Nucleotide diversity (π) with the Jukes and Cantor (1969) correction taking into account the average number of nucleotide substitutions per site was calculated together with Θ per site (Watterson 1975) derived from the number of polymorphic sites (S). We also calculated the number of haplotypes (H) together with haplotype diversity (H_d; Nei and Tajima 1983) and the average of nucleotide differences (K). To gain insights to evolutionary forces acting on each region we calculated a separate overall statistics for indel polymorphisms. This assumption fitted well with the binary indel data coded by SeqState, since DnaSP also uses the SIC method. Based on indel information the total number of indel events (I), average indel length (L), number of indel haplotypes (H_i) together with diversity values such as indel

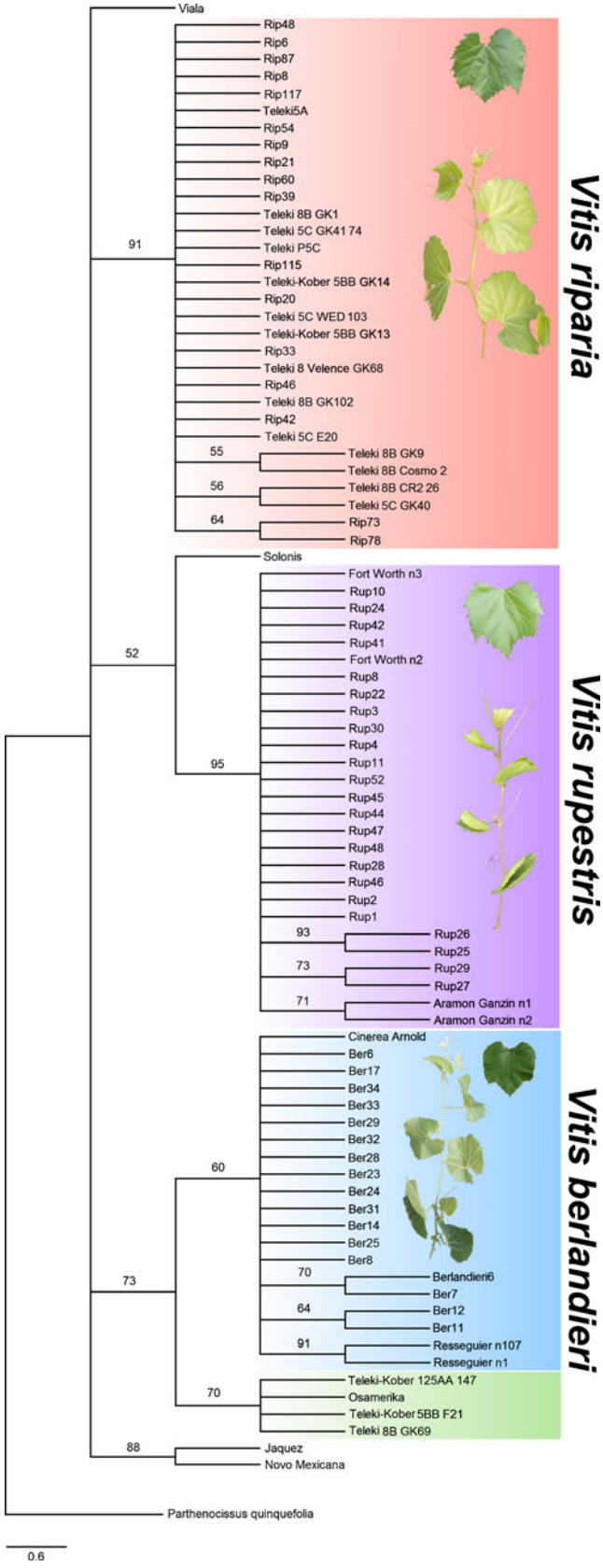

Fig. 3 Maximum likelihood (ML) majority rule consensus tree for the analysed *Vitis* accessions and outgroup taxon obtained from plastid DNA data. *Red shading V. riparia* clade, *purple shading V. rupestris* clade, *blue shading V. berlanidari* clade. The group of Teleki accessions appearing as sister group in the *V. berlandieri* clade is shaded in *green*. Images of representative leaves and grooving shoots with tendrils are shown

haplotypes diversity (H_{id}), indel diversity (K_i) and indel diversity per site (π_i) was calculated. The D test statistic proposed by Tajima (1989) was calculated to test the hypothesis that all indels are selectively neutral.

Results

Genetic Relationships and Phylogenetic Inference

All parsimony analyses resulted in a set of equally parsimonious trees (EPTs) with the length of 183 steps. The ensemble consensus index (CI; Kluge and Farris 1969) and ensemble retention index (RI; Farris 1989) were 0.74 and 0.89, respectively. The (strict) consensus trees of different analyses were identical, as illustrated in Fig. 2. Because of these highly identical results we did not even attempt to find all EPTs. This would be unnecessary and simply waste of CPU time as noted already by Farris et al. (1996). The ML analysis resulted in a tree (Fig. 3) with a log likelihood score of 4,232.1161. This tree was consistent with the tree obtained in the parsimony analysis. The results of the Bayesian runs with different alignments were highly congruent with each other. PSRF values averaged 1.0000–1.0001 strongly suggesting that stationarity had been reached. The comparison of the topologies and associated posterior probability values obtained across the independent runs for each alignment with AWTY verified convergence. The inclusion of binary indel characters in the alignments influenced the convergence of the runs. Runs without indels reached stationarity after 8×10^6 generations while the inclusion of binaries shortened this to 6×10^6 generations. The resulting trees from the analyses of different datasets did not show any incongruence, but the inclusion of indel characters produced more resolution in the deeper nodes. The majority rule consensus tree is illustrated in Fig. 4, complemented with the support values gained for each dataset.

When the consensus trees obtained from different analyses are compared it is obvious that the concatenated dataset also includes characters that are congruent with other characters. Both analyses (parsimony and BI) that included also the binary indel data resulted in a consensus trees with more resolution than the trees of the ML analysis performed

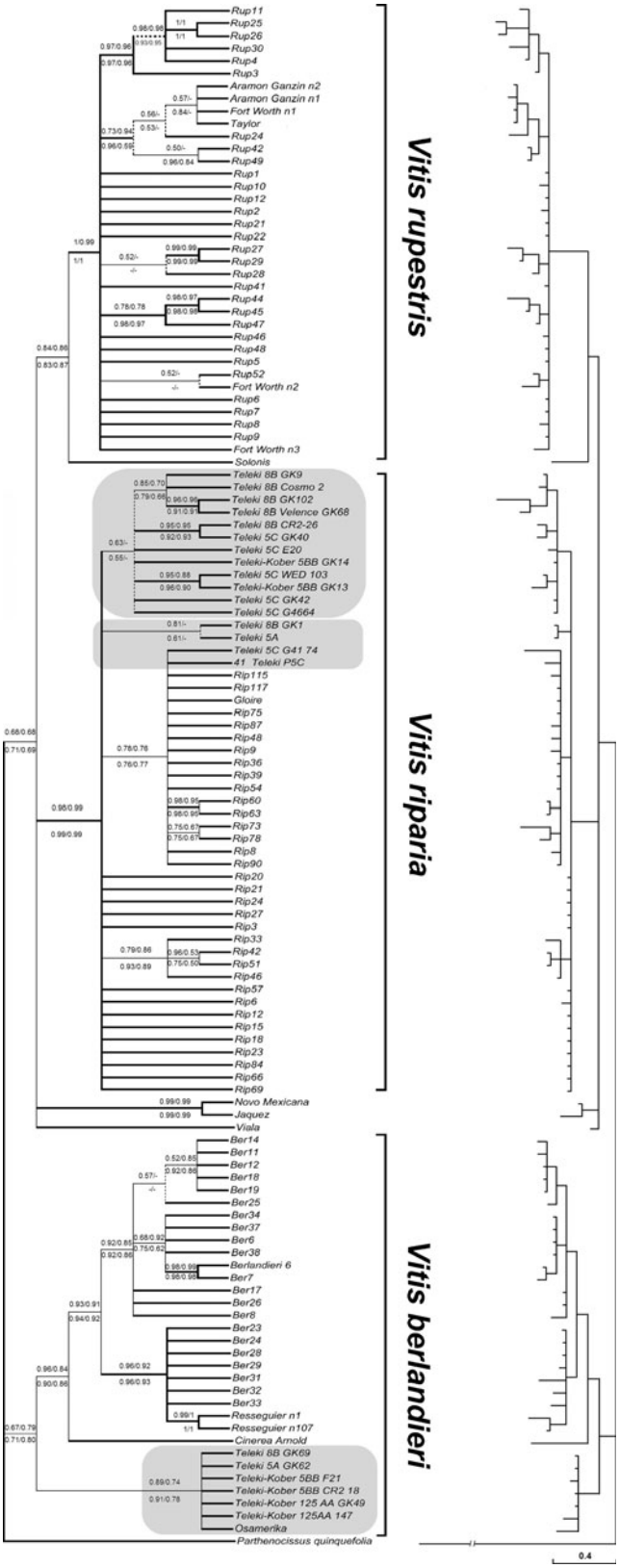

Fig. 4 Bayesian 50 % majority rule consensus tree (*left*) and phylogram (*right*) from the analysis of the concatenated dataset of chloroplast regions. Numbers above the branches are posterior probabilities (PP); the Gblocks + indels are shown on the left, and Gblocks alignment without indels is shown on the right. Below the branches, PP values for MUSCLE + indels are shown on the left and MUSCLE alignment without indels on the right. Resolution with high (PP 95 %) values present in at least two of the alignments is indicated by *bold lines* while *dashed lines* specify topologies that collapsed or were absent in the results of at least one alignment. *Grey shaded boxes* represent the analyzed Teleki lines on the consensus tree

without indel data. The nodes with high support values were taxonomically clearly patterned, suggesting that the analyzed *Vitis* species are well differentiated. This provides possibility to trace the maternal ancestry of the Teleki lines, which seem to group in two clades. Six Teleki lines were placed within the *V. berlandieri* clade. Two accessions (41 Teleki P5C and Teleki 5C G41 74) were nested within a bigger *V. riparia* clade formed by many *V. riparia* accessions; however, with only moderate PP support values. Two other Teleki lines (Teleki 8B GK1 and Teleki 5A) were grouped together, while the rest of the Teleki accessions were united in a separate clade. Both of these groups appeared only with the inclusion of the binary indel dataset and got only weak support values, indicating that only indel characters are responsible for the formation of these topological arrangements. The pure *V. rupestris* selection Fort Worth 1, 2 and 3 formed a group in the *V. rupestris* clade with low PP support value. A group was formed by Aramon Ganzin 1 and 2, Fort Worth 1 and Taylor. The first two are known as *V. vinifera* × *V. rupestris* with *V. rupestris* as a female parent. The PP support value of this group was very low. Taylor has been reported as a *V. riparia* × *V. labrusca* L. hybrid, but we have it in the *V. rupestris* clade. *V. solonis* Prince ex Bailey is an interspecific hybrid between *V. riparia* and *V. mustangensis* Buckley (Syn.: *V. candicans* Engelm. ex Durand) and it was resolved between the *V. rupestris* and *V. riparia* clades. Riparia Gloire de Montpellier is found in the *V. riparia* group with relatively high PP support value as a pure *V. riparia* and it shows close relationship with Teleki 5C G41 and Teleki P5C, which could indicate parent relationships. *Vitis novo-mexicana* Lemmon ex Munson and Jaquez formed a clade next to the *V. riparia* clade, with very high (0.99) PP support value, indicating a highly similar genetic background. Their parentage has been reported so far as *V. riparia* × *V. mustangensis* and *V. bourquina* (*V. vinifera* × *V. aestivalis* Michx.), respectively (Galet 1967). Vialla has been claimed to have *V. labrusca* and *V. riparia* background (Galet 1967), this is in accordance with our results. Resseguier N1 and N107 are very similar, form a clade and fit well into the *V. berlandieri* clade as expected because they are pure *V. berlandieri* selections. They were possibly one of the parents of the Teleki seedlings as most of the canes originated from the seeds obtained from Mr. Resseguier and his plantations. A group of Teleki seedlings are genetically close to that group. It seems that Teleki-Kober 5BB and 125 AA could have a *V. berlandieri* maternal background.

Variability of the Three Chloroplast Regions

All datasets obtained for the complete *trn*L-*trn*F and *trn*S-*trn*G regions showed variation either in their length or in their nucleotide composition. For the analyzed terminals the aligned data matrix, consisting of the *trn*L group I intron (542–543 bp), the *trn*L-*trn*F spacer (431–457 bp) and the *trn*S-*trn*G spacer (1,021–1,076 bp), included 2,159 characters with gaps. Sequence statistics were calculated for the concatenated MUSCLE alignment, distinguishing the features of the two spacers and the group I intron which are summarized in Table 1. The Gblocks alignment consisted 1,913 bp representing 88 % of the original 2,159 bp long MUSCLE concatenated alignment. This final alignment was formed by 23 selected blocks with flanking positions allowing minimum length of 10 bp for each block, while the minimum number of sequences for a flanking position was 104. The two alignments are provided as Supplementary Tables 2 and 3. This allowed elimination of divergent, poorly aligned regions that may not be homologous and that may have been saturated by multiple substitutions. The *trn*L intron (5.51 %) and the *trn*L–*trn*F spacer (5.58 %) contained variable sites at practically equal levels, but length variation was more common in the non-coding spacer region. The other plastid region *trn*S-*trn*G, showed slightly higher number of variable sites (6.83 %) and lower level of conserved positions. In the complete *trn*L-*trn*F region only 18 of the total 57 variable characters turned out to be parsimony informative this being higher in *trn*S-*trn*G with 76 variable characters and 28 parsimony informative sites. Both spacers (*trn*L-*trn*F: 35.74 %; *trn*S-*trn*G: 34.85 %) and the intron region had similar GC content (*trn*L intron: 39.44 %). The ratio of transition to transversion (ti:tv) ranged from 0.27 to 0.45 in the *trn*L intron and *trn*L-*trn*F spacer, respectively. The *trn*S-*trn*G spacer had a slightly higher ti:tv value of 0.66, while the overall value was 0.57 for the concatenated dataset. The ratio showed no bias towards either transitional or transversional substitution because the two kinds or substitutions were almost equally probable.

The genetic diversity indices calculated for each species, chloroplast region and for Teleki line groups based on their putative ancestry inferred by the phylogenetic analysis are presented in Table 2. These indices were also calculated for all Teleki lines and terminals. The nucleotide diversity values π (JC) and Θ (S) for the spacers ranged between 0–0.0031 and 0–0.0068, while they were slightly lower for the intron, varying between 0.0016–0.0021 and 0.0019–0.0052, respectively. The trend was similar in the average differences between pairs of nucleotides (K) among groups ranging 0–3.108 for the spacers, and 0.571–1.068 for the intron. The occurrence of substitutions along the analyzed concatenated sequences is shown in Fig. 5. Within wild grapevine species the inferred number of possible haplotypes was highest for *Vitis rupestris* (25) and lowest for *V. riparia* (14). The haplotype diversity—the probability that two randomly selected haplotypes are present in the sample—based on the total dataset varied between 0.714 and 0.960. All these calculated indices of diversity confirm that variation

Table 1 Sequence statistics for the two non-coding and the intron based on the total aligned data set. *SE* Standard error (calculated based on 100 bootstrap replicates)

	*trn*L intron	*trn*L-*trn*F spacer	*trn*L-*trn*F Complete	*trn*S-*trn*G spacer	Combined
Sequence length range, including hot sport (bp)	542–543	431–457	973–1,000	1,021–1,076	2,021–2,053
Average length (SD)	542 (0.49)	432 (2.82)	974 (2.83)	1,072 (4.71)	2,046 (3.03)
Aligned length	544	502	1,046	1,113	2,159
Variable characters (%)	30 (5.51)	27 (5.58)	57 (5.45)	76 (6.83)	133 (6.16)
PI characters[a]	10	6	18	28	46
Conserved sites	513	414	928	999	1,927
GC content (%)	39.44	35.74	35.41	34.85	31.51
ti : tv[b] (SE)	0.27 (0.21)	0.45 (0.31)	0.31 (0.19)	0.66 (0.4)	0.57 (0.28)

[a] Parsimony informative (PI) characters

[b] Transition/transversion ratio

was low within each wild grapevine species represented by the assembled accessions, and further indicates that each group was well differentiated as all values were higher when calculated among all accessions. This was clearly apparent for the two groups of Teleki lines, which exhibited lower diversity when treated separately and higher values when treated together.

Microstructural Changes and Indel Variability

A number of length mutations were identified in all plastid DNA regions used. These changes yielded additional 30 binary characters coded for 26 microstructural changes. The list of indels and their coded states in binary format is given in Supplementary Table 4. In the *trn*S-*trn*G spacer most of these indels represent simple sequence repeats (SSR), mostly duplications of tri- to hexa-nucleotide repeats or different stretches of poly-A mononucleotide repeats. An exception is a larger (77 bp) insertion restricted to the outgroup *Parthenocissus quinquefolia* (see Supplementary Table 2), which seemed to be lost in *Vitis*. *Parthenocissus* sequence seemed to be the most variable in length, accumulating larger insertions/deletions. The *trn*L intron included only a few one base (A/T or C) indels, while the *trn*L-*trn*F spacer had numerous poly-T stretches of varying length. Such T-rich regions downstream of plastid tRNA genes are typical regulatory elements for the termination of transcription together with other hairpin structures (Won and Renner 2005). Separate diversity statistics were calculated only for the indels and the results are presented in Table 3. These indels were excluded from nucleotide based diversity indices and were summarized for all accessions. The reason for this is that we were interested in the overall diversity of these indels and their distribution within intron and spacer regions, and their contribution to improving the resolution and to differentiating the analysed terminals. Based on these statistics the average length of indels was much longer in the spacers—16.17 in the *trn*L-*trn*F and 3.93 in the *trn*S-*trn*G—while it was only one base pair in the *trn*L intron. According to other diversity measures related to indel events, such as the indel diversity (K_i) and indel diversity per site (π_i), together with Θ_i were highest in the *trn*S-*trn*G spacer (K_i= 1; π_i=0.0095; Θ_i=2.785), smaller in the *trn*L-*trn*F spacer (K_i=0.885; π_i=0.0023; Θ_i=1.857) and lowest in the *trn*L intron (K_i=0.653; π_i=0.0013; Θ_i=0.743). This suggests that indel characters exhibit additional diversity at a level comparable to that of nucleotide diversity. We used Tajima's *D* statistics to test whether the distribution of indels differs from the expected under the null neutral model and thus signifies the presence of selection for indels. These tests showed no significance (P>0.10) ,indicating that both indel diversity and diversity per site can be assumed to be generated by the same process. Negative *D* values might indicate a recent effect of positive selection increasing the frequency of indels mutations.

Discussion

Mixed Origin of the Teleki Lines

The genetic diversity of grapevine rootstocks has been treated in a number of studies (Jahnke et al. 2009; Carimi et al. 2011). These have shown clearly that information of the genetic diversity of the rootstocks can be used successfully for efficient management of *Vitis* germplasm and genotype selection for different purposes. Teleki lines have been used widely because of their favorable traits but so far the relationships of these rootstocks have been obscure. Their unclear origin (as described in the Introduction) makes detailed studies necessary. We found great differences within the so-called 'clones' of Teleki 5C and Teleki-Kober (TK) 5BB. Suggested clones of Teleki 5C had both indel and sequence polymorphism, while TK 5BB 'clones' showed even greater

Table 2 Diversity statistics calculated within each wild species and among each group. Indices are given for separate plastid regions. *N* Number of accessions, *S* number of polymorphic sites, *H* number of haplotypes, π (JC) nucleotide diversity, H_d haplotype diversity, *K* average of pairwise nucleotide differences

	*trn*L	*trn*L-*trn*F	Combined *trn*L-*trn*F	*trn*S-*trn*G	Total[a]
Vitis riparia					
N	34	34	34	34	34
S	5	3	8	9	17
π (JC)	0.0017	0.0007	0.0012	0.0007	0.0009
Θ *(S)*	0.0024 (0.0012)	0.0023 (0.0014)	0.0022 (0.0010)	0.0022 (0.0095)	0.0020
H	6	4	10	7	14
H_d	0.656 (0.059)	0.223 (0.093)	0.720 (0.064)	0.417 (0.105)	0.820 (0.055)
K	0.859	0.232	1.091	0.740	1.831
Vitis rupestris					
N	36	36	36	36	36
S	4	8	12	8	21
π (JC)	0.0021	0.0028	0.0022	0.0019	0.0019
Θ *(S)*	0.0019 (0.0011)	0.0068 (0.0029)	0.0036 (0.0014)	0.0019 (0.0009)	0.0025 (0.0009)
H	5	8	13	12	25
H_d	0.619 (0.076)	0.516 (0.099)	0.805 (0.060)	0.825 (0.042)	0.960 (0.020)
K	1.068	0.900	1.968	1.868	3.837
Vitis berlandieri					
N	24	24	24	24	24
S	10	6	16	9	26
π (JC)	0.0020	0.0015	0.0018	0.0025	0.0020
Θ *(S)*	0.0052 (0.0022)	0.0050 (0.0025)	0.0051 (0.0020)	0.0024 (0.0011)	0.0034 (0.0012)
H	7	4	8	14	17
H_d	0.430 (0.124)	0.230 (0.110)	0.490 (0.123)	0.917 (0.037)	0.943 (0.031)
K	1.020	0.480	1.580	2.543	4.123
Teleki (Riparia)					
N	16	16	16	16	16
S	4	0	4	10	22
π (JC)	0.0016	0	0.0009	0.0031	0.0019
Θ *(S)*	0.0024 (0.0014)	0	0.0014 (0.0008)	0.0055 (0.0023)	0.0033 (0.0013)
H	5	0	5	10	13
H_d	0.600 (0.127)	0	0.600 (0.127)	0.867 (0.079)	0.950 (0.048)
K	0.808	0	0.808	3.108	3.917
Teleki (Berlandieri)					
N	7	7	7	7	7
S	2	0	2	1	3
π (JC)	0.0020	0	0.0006	0.0003	0.0004
Θ *(S)*	0.0052 (0.0022)	0	0.0009 (0.0007)	0.0004 (0.0004)	0.0006 (0.0004)
H	3	0	3	2	4
H_d	0.430 (0.124)	0	0.430 (0.124)	0.286 (0.196)	0.714 (0.181)
K	0.571	0	0.571	0.286	1.687
Teleki (All)					
N	23	23	23	23	23
S	5	0	5	18	23
π (JC)	0.0014	0	0.0008	0.0042	0.0024
Θ *(S)*	0.0027 (0.0014)	0	0.0015 (0.0008)	0.0049 (0.0019)	0.0031 (0.0012)
H	6	0	6	11	16
H_d	0.569 (0.114)	0	0.569 (0.114)	0.874 (0.048)	0.949 (0.031)

Table 2 (continued)

	*trn*L	*trn*L-*trn*F	Combined *trn*L-*trn*F	*trn*S-*trn*G	Total[a]
K	0.735	0	0.735	4.221	4.957
Among all acc.[b]					
N	122	122	122	122	122
S	28	16	41	39	81
π (JC)	0.0028	0.0015	0.0020	0.0062	0.0041
Θ *(S)*	0.0105 (0.0030)	0.0095 (0.0031)	0.0088 (0.0024)	0.0074 (0.0020)	0.0075 (0.0019)
H	22	14	34	45	71
H_d	0.711 (0.037)	0.300 (0.055)	0.774 (0.034)	0.912 (0.019)	0.969 (0.008)
K	1.312	0.468	1.740	6.185	8.344

[a] Indices calculated for the complete concatenated alignment including partial or complete exon regions, but excluding gaps

[b] Indices calculated excluding the outgroup (*Parthenocissus quinquefolia*)

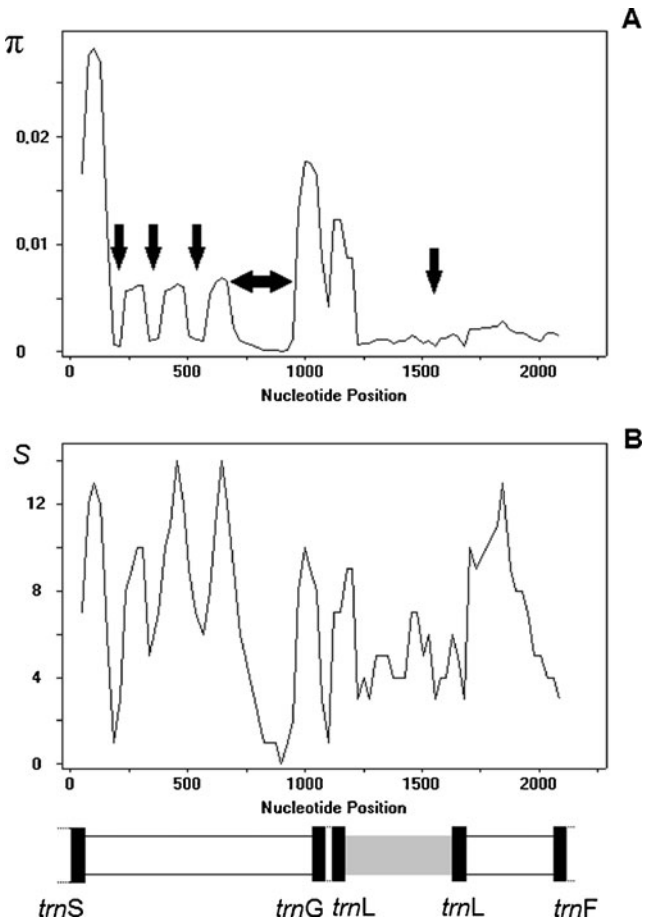

Fig. 5a,b Hundred base-pair sliding windows with 25-bp step size of the concatenated chloroplast DNA dataset of the *trn*S-trnG and complete *trn*L-*trn*F regions of the analyzed grape accessions. The windows highlight the low level of nucleotide diversity along the *trn*L-*trn*F plastid DNA region compared to the *trn*S-*trn*G region. *Black bars* Partial and complete exons of tRNA genes; *white boxes* non-coding spacer regions, *grey box* intron sequence. **a** Variability of nucleotide diversity π (JC), **b** number of polymorphic sites (*S*). *Arrows* indicate footprints of selection

variation, placing some terminals in the *V. berlandieri* clade in all performed phylogenetic analysis, indicating that these putative clones are not identical. This confirm the results of Podmaniczky et al. (2006), who suggested that the selection of Teleki 5C was performed on heterogenous material collectively termed as 'Teleki 5C', or misidentification based on morphological characters. Putative Teleki-Kober clones also originate from mixed source of plant material sent by Z. Teleki to F. Kober, who made further selections based on this material. TK 5BB accessions formed a sub-clade with Teleki lines 125 AA in the BI with 0.74–0.91 PP support depending on the inclusion of indel binaries. This topology was also consistently present in the parsimony and ML analyses (in the former with jackknife support value <50 % and in the latter with only moderate 70 % bootstrap support). These accessions may originate from *V. berlandieri* female parents. However, this grouping is supported by PP values of only 0.67–0.80 in the BI depending on the

Table 3 Overall diversity statistics for indels. *I* Total number of indel events analyzed; *L* average indel length; H_i number of indel haplotypes; H_{id} indel haplotype diversity; K_i indel diversity; π_i indel diversity per site; Θ_i, indel diversity from *I*. Tajima's *D* tests were not significant ($P>0.10$)

Indels	*trn*L	*trn*L-*trn*F	Combined *trn*L-*trn*F	*trn*S-*trn*G	Total[a]
N	123	123	123	123	123
I	4	11	15	15	30
L	1	16.17	12,03	3.93	6.875
H_i	5	8	14	10	26
H_{id}	0.568	0.641	0.821	0.600	0.884
K_i	0.653	0.885	1.884	1	2.884
π_i	0.0013	0.0023	0.0020	0.0095	0.0013
Θ_i	0.743	1.857	2.785	2.785	5.571
Tajima's *D*	−0.2306	−1.2992	−0.8720	−1.7278	−1.4339

[a] Indices calculated including the outgroup (*Parthenocissus quinquefolia*)

alignment, and 73 % bootstrap value in the ML analysis. This sub-clade forms a sister group to the rest of the *V. berlandieri* accessions and is consistently present in all performed analyses. Using *trn*L-F sequences for the phylogenetic reconstruction of the genus *Vitis*, Tröndle et al. (2010) obtained also trees with large polytomies when the *trn*L-F region alone was used. The resolution of the deeper nodes increased when other chloroplast regions (*trn*L and *trn*K-*mat*K intron) were added to the analysis. The major backbone clades of the tree remained but with increased support values, but some deeper relationships were weakly resolved even with the addition of further non-coding cpDNA markers. Phylogenetic analysis of the whole family Vitaceae by Ren et al. (2011) included chloroplast markers like *trn*C-*pet*N, *trn*H-*psb*A in addition to *trn*L-F and showed that non-coding sequence diversity is variable also within the genus *Vitis*. This variation should be taken into account when closely related species such as North American *Vitis* taxa are investigated. We observed high level of sequence conservation among the analyzed accessions. This contrasts with the results obtained for the genus *Hordeum*, where different species of the same genus or, even at infraspecific level, unique haplotypes can be found with these kinds of sequences (Provan et al. 1999). If the Teleki accessions forming a sister group to the Berlandieri-clade possibly originate from *V. berlandieri* female parents, they were selected from other haplotypes not sampled in our analysis. The close relationship of these accessions with the Berlandieri-clade has moderate support values, but it seems clear that they do not share the same ancestry with other accessions of Teleki rootstocks resolved in the *V. riparia* clade with high support value. To gain further resolution it would be desirable to include additional non-coding sequences or even to combine our approach with the use of data obtained from other DNA fingerprinting techniques used successfully with other species (e.g. Carvalho et al. 2012; Ovesná et al. 2012; Mandoulakani et al. 2012; Wang et al. 2012). Nuclear markers analyzed together with the cpDNA markers would be useful to explore the recent history of Teleki rootstocks and may assist in tracing their origin more precisely.

The origin of the rootstocks indicates which species have contributed to their characters and information about important cultural traits can be found for further breeding development. The Teleki rootstocks are in different groups of resistance against root-knot nematodes. Among the analyzed terminals, 5BB and 8B have been reported to be resistant (Snyder 1936; Boubals 1954, 1979; Sauer 1967; Bouquet and Dalmasso 1976; Santibanez 1983). This differs what is known about resistance against the northern root-knot nematode (*Meloidogyne hapla* Chitwood, 1949). *Vitis riparia* and *V. rupestris* rootstocks are the most resistant, followed by 5BB. These rootstocks differ also in their resistance against California dagger nematode (*Xiphinema index* Thorne and Allen, 1950). The Teleki 5C has been found to be sensitive (Harris 1983; Jin 1997). Lime tolerance also differs, the Galet-index of 5BB 20 is higher than that of the other rootstocks (Galet 1947). In very dry soils, *V. riparia*-*V. berlandieri* crossings are in the order of diminishing resistance of 8B, 5BB. In addition, utilisation of rootstocks with diverse genetic background has shown differences, for example in adaptation to different soil conditions. Nowadays "fine tuning" has been initiated in the usage of rootstocks in order to find the best combinations for a given site in order to improve grape and wine quality (Kocsis and Lehoczky 2000; Renouf et al. 2010).

It is important to note that Teleki 8B may also originate from mixed sources, as explained for Teleki 5C and TK 5BB above. However, only one accession was placed in the *Vitis berlandieri* clade. According to our results, Teleki 8B and Teleki 5C are close relatives, and possibly originate from *V. riparia* female parents. This is reflected in their characters such as ease of rooting and grafting, and only moderate or weak tolerance of drought. Jahnke et al. (2012) in their latest study based on microsatellite analysis suggested parent–offspring combinations for Teleki rootstocks T5C GK42, T5C E20, T5C GK46, and T5C WED. They assume that the possible parent could have been *V. riparia* 'Gloire de Montpellier' rootstock. Based on the cpDNA markers, it seems likely that the maternal parent of these Teleki lines, together with others grouped in the *V. riparia* clade, is undoubtedly a male flowered *V. riparia*. However, due to small differences in the cpDNA regions analysed, we do not as yet have enough information to clarify this. The Teleki and Gloire sub-clades within the *V. riparia* group had only low to moderate PP values, with only minor microstructural changes observed in their sequences. However, the entire *V. riparia* clade was also found in ML (bootstrap value 91 %) and parsimony (jackknife value 75 %) analyses. The exclusion of indels from the analyses would have resulted in the *V. riparia* clade as a large polytomic group with only some groups of two terminals. This agrees well with the ML analysis where these groups were resolved in a large polytomy also suggesting that indel characters play an important role in sub-clade formation. Supplementary nuclear genomic information will be needed to trace the complex history of crossings.

Implications for Grapevine Breeding

Combining our maternally inherited cpDNA data with nuclear SSR markers could reveal both paternal partners for these putative clones. In grape breeding, this work has been initiated based on molecular markers (Bowers et al. 1999b) in combination with pedigree analysis (Vouillamoz et al. 2004). This has been done in order to improve germplasm

utilisation, and to further facilitate the selection of parents in the development of rootstock lines with various traits of interest (Lowe and Walker 2006). Great efforts have been made in recent years to develop several sets of simple sequence repeat (SSR) markers for the genetic analysis of the grapevine genome. The main rationale behind this was to compile and develop a standard set of molecular tools and a database (e.g. the Grape Microsatellite Collection, http://meteo.iasma.it/genetica/gmc.html) containing reference microsatellite profiles for true-to-type identification of grapevine accessions (This et al. 2004). With other cultivated plants, such as tomato (Bredemeijer et al. 2002) and wheat (Röder et al. 2002), such databases have led to outbursts of studies in many laboratories. In complicated cases such as grape, where for some cultivars many synonyms and numerous homonyms exist, coupled with the uncertainty about the origin of germplasm, molecular markers will certainly be beneficial. Our data indicate that the non-coding cpDNA regions represent a good source of genetic characters for grapevine studies. In addition, these plastid markers allowed us to infer the possible maternal ancestral origins of the breeding material. Such non-coding plastid sequences have been used to study the phylogeny of the genus *Vitis* (Tröndle et al. 2010) and Vitaceae (Soejima and Wen 2006; Ren et al. 2011), and are applied widely in DNA barcoding (Taberlet et al. 2007). These regions, despite the low level of nucleotide diversity detected, provided useful measures of haplotype diversity and clear information on genetic relationships. This is in agreement with the studies of *Allium roseum* L. (Guetat et al. 2010), *Festuca arundinacea* Schreb., *F. pratensis* Huds. (Hand et al. 2012) and *Panicum virgatum* L. (Zalapa et al. 2011). However, plastid single nucleotide polymorphisms (SNPs) may not be able to discriminate all the accessions analysed, but they are potential candidates for phylogenetic analyses of the cultivars. Recent studies have also shown that SNPs would be highly useful in grapevine research offering possibilities of automation with highly repeatable results based on the partial or complete genome sequencing identifying thousands of SNPs (Cabezas et al. 2011).

Genetic Diversity and Species Divergence

North American *Vitis* taxa belonging to subgenus *Vitis* are considered as a complex of ecospecies that are genetically close despite their phenotypic variation (Levadoux et al. 1962). Genetic diversity analysis confirms these findings since all cpDNA regions showed low level of sequence polymorphism, which may result from a low rate of molecular evolution. The outgroup taxa *Parthenocissus quinquefolia* showed many autapomorphic microstructural changes and sequence positions. The clear separation also at the molecular level could suggest separation of *Parthenocissus* from *Vitis* early in Vitaceae (Péros et al. 2011). Our genetic diversity analysis coupled with the phylogenetic analyses separated the analyzed taxa clearly into three major clades, suggesting that gene flow between *V. riparia*, *V. rupestris* and *V. berlandieri* must be restricted. However, these species are often sympatric and occur together in the same habitat. Several authors have pointed out, that members of subgenus *Vitis* easily cross with each other and there are no significant breeding barriers, and this has lead to multiple hybrids and confused taxonomy (Aradhya et al. 2008). Genetic diversity data shows that within species diversity is low, undoubtedly attributed to selection for breeding purposes and this can be seen in the distribution if variable sites in the cpDNA regions analyzed (see Fig. 5). Genetic diversity preserved between species was clearly higher than within species (Table 2), indicating that reproductive isolation have evolved among the analyzed taxa with overlapping distributions, e.g., *V. riparia* and *V. berlandieri*. Such isolation might be due to the differences in flowering time. For example *V. riparia* is known flower earlier than *V. berlandieri* (Levadoux et al. 1962). These results confirm the recent findings of Zecca et al. (2012) who made the same conclusions using phylogenetic analysis. Schmidt et al. (2009) also mention that difference in flowering time in *V. berlandieri* and *V. riparia* is responsible for the maintenance of pure species populations with minimal interspecific hybridization in nature. This distinction cannot be made for *V. cinerea* and *V. berlandieri* in our analysis. However, this should be treated cautiously because only one exemplar of *V. cinerea* was included in our study. In the BI and parsimony analyses, the sole *V. cinerea* terminal was a sister group to *V. berlandieri* accessions. However, this topology was not present in the ML tree. Moor (1991) treated *V. berlandieri* as *V. cinerea* var. *helleri* based on the fact that 'var. *helleri*' is the oldest varietal name attributed to this taxon and must be used if this taxon is recognised at the varietal level. Our present molecular evidence supports this, but more sampling is needed to confirm this. Although genetic diversity indices showed more or less the same diversity within the analyzed species, *V. berlandieri* showed the highest average of pairwise nucleotide differences (K= 4.123). Since molecular divergences usually precede population divergence a high number of nucleotide differences in pairwise sequence comparisons can indicate a higher upper boundary for population divergence times.

Utility of Indels: Exclude, Include or Dismiss

The increasing number of complete chloroplast genomes available for several angiosperms makes it evident that they contain many repeated sequences in addition to the major inverted repeat (IR; Saski et al. 2005). The analysis of non-coding regions has many practical implications for

germplasm assessment relying on the conserved nature of the chloroplast genome. The non-coding spacer regions are probably the tool most widely used in plant science to study a number of phylogenetic and molecular ecology problems due to the low rate of evolution and absence (or nearly so) of recombination in the plastome (Rossetto et al. 2001). Universal PCR primers can be constructed easily on the basis of conserved coding sequences of cpDNA genes to amplify products for further downstream applications, direct sequencing and further development of other useful molecular tools based on the structural elements. Closer inspection of *Vitis* plastid sequences in our analyses revealed the abundance of many different microsatellite regions, distributed mostly in the analyzed non-coding regions of *trn*L-*trn*F and *trn*S-*trn*G. These SSRs were frequently mononucleotide repeats of poly-A or -T stretches that were previously reported to be common and abundant in chloroplast genomes of polysporangiophytes, since the AT content of the plastome is much higher than in the nuclear genome (Wakasugi et al. 1998; Ishii and McCouch 2000; Ogihara et al. 2000; Khan et al. 2012). The amount of variable positions was lowest in the *trn*L intron and higher in the two spacer regions. This might be due to the fact that the *trn*L region containing the sole group I intron in the chloroplast genome has to maintain the secondary structure and critical sequence elements for its self-splicing from precursor-RNAs (Kuhsel et al. 1990; Simon et al. 2003). The slight difference between the ti:tv ratio of the intron and spacers shows that the intron evolves somewhat slower than the spacers. Previous studies have suggested that this sole group I intron of the chloroplast genome has identical evolution rate with the protein coding genes (Zurawski and Clegg 1984). It has been shown in many plant groups that non-coding regions often incorporate polynucleotide stretches (e.g. poly-A and -T) consisting mostly of microsatellites and different sequence arrangements such as inversions (Borsch et al. 2003). The flanking regions of these short stretches of SSRs present in cpDNA have been shown to be orthologous in multiple alignments, and this has enabled the construction of universal primer set cpSSRs for a variety of taxa (Weising and Gradner 1999). Both the phylogenetic and indel diversity analyses indicated that the indels contain information to differentiate *Vitis* species and rootstock lines. The complete sequence of the grapevine chloroplast genome revealed further 36 direct and inverted repeats found mostly in intergenic spacer regions (Jansen et al. 2006). These repeats might be used in the future breeding research. Arroyo-García et al. (2002) already found intra- and inter-specific cpSSR length variation in *Vitis*. The development of SSRs for non-coding chloroplast regions has provided higher resolution for the study of closely related species. Nuclear microsatellites have not been so useful to analyze genetic relationships of species in many plant genera (Ishii and McCouch 2000), while in *Vitis* they have provided better resolution for closely related cultivars (Arroyo-García et al. 2002). These findings are in accordance with our observations about the inclusion of indel binary data present in spacer loci. With the closely related accessions, the inclusion of binary data in the phylogenetic analysis slightly improved resolution in deeper nodes, albeit with weak support values (Fig. 4). An example for the increase of resolution can be seen from the Aramon Ganzin n˚1, n˚2 and Fort Worth n˚1 plus Taylor group, which received 0.57 (Gblocks + indel) and 0.84 (MUSCLE + indel) PP values in different alignments but were absent in the Bayesian analysis without indel binary characters. This group was also resolved as a large polytomy in parsimony analysis while the ML without indel characters grouped only the two Aramon Ganzin accessions together with 71 % bootstrap support value. The presence of this topology in the BI was due to a specific microstructural change $(TATCC)_n$ in the *trn*L-*trn*F spacer. This pentanucleotide repeat was present from aligned base 1,736 to 1,750 and has an additional $(TATCC)_2$ copy in these accessions. The use of indel characters can lead to increased resolution, and this supports utilisation and inclusion of cpSSR changes in further analysis of the *Vitis* gene pool. We recommend use of methods and software allowing concatenated matrices and mixed sets. This can also be seen from our experiments with the two different alignment methods Gblocks and MUSCLE. The former employs a user defined conservative pruning technique on the alignment, finding the blocks of interest then removing the non-conserved positions (mostly associated with gaps) from the alignment. The progressive MUSCLE alignment is based on fast distance estimation, log-expectation scoring and refinement with tree-dependent restricted partitioning that tries accurately to place gaps in the alignment. Gaps in alignments might be congruent with other characters as can be seen in our results. Overall, indels do not contradict other characters and thus their exclusion should be considered only in those cases where ambiguity clearly increases. This was reported to happen with the increase of genetic and taxonomic distance by Nichols and Barnes (2005) and Dunn et al. (2008).

Concluding Remarks

In conclusion, the use of two non-coding and one group I intron plastid sequence allowed analyses of the diversity and relationships of grapevine rootstock material. The indel characters found in these plastid sequences also proved to be useful. Genetic differentiation analysis of the accessions provided prospects for identifying diverse parental combinations for the widespread Teleki rootstock lines. Our results could aid in segregating populations for genetic

studies, and help to identify germplasm material for the introgression of desirable genes from diverse sources. However, analyses of a more extensive collection of the rootstock germplasm is desirable to define the genetically more closely related core set of putative maternal parents of Teleki lines to further refine their relationships and placement in the *Vitis* gene-pool. Given the strict maternal inheritance of plastid DNA in *Vitis*, analyses of the nuclear genome should also be considered in future studies to track the paternal ancestry of the rootstocks. We are initiating such studies based on microsatellite markers and nuclear DNA sequences to provide further insights to the origin of the economically important Teleki lines.

Acknowledgements Thanks are due to Kinga Klára Mátyás for laboratory assistance, Nándorné Nagy and Tamás Lönhard for field collections. We also thank Ildikó Varga for her help preparing the figures. This work was supported by the Hungarian Scientific Research Fund (project no. K73708). Peter Poczai gratefully acknowledges support from the Mare Curie Fellowship Grant (PIEF-GA-2011-300186) under the seventh framework program of the European Union.

References

Adam-Blondon AF, Roux C, Claux D, Butterlin G, Merdinoglu D, This P (2004) Mapping 245 SSR markers on the *Vitis vinifera* genome: a tool for grape genetics. Theor Appl Genet 109:1448–1458

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19:716–723

Aradhya M, Koehmstedt A, Prins BH, Dangl GS, Stover E (2008) Genetic structure, differentiation, and phylogeny of the genus *Vitis*: implications for genetic conservation. Acta Hortic 799:43–49

Arroyo-García R, Lefort F, de Andrés MR, Ibánez J, Borrego J, Jouve N, Cabello F, Martínez-Zapater JM (2002) Chloroplast microsatellite polymorphisms in *Vitis* species. Genome 45:1142–1149

Bakonyi K, Kocsis L (2004) Teleki Zsigmond élete és munkássága. University of Veszprém, Keszthely, p 64

Borsch T, Hilu KW, Quandt D, Wilde V, Neinhuis C, Barthlott W (2003) Noncoding plastid *trn*T-*trn*F sequences reveal a well resolved phylogeny of basal angiosperms. J Mol Evol 16:558–576

Boubals D (1954) Les nematodes parasites de la vigne. Progr Agr Et Vit 141(173–182):204–207

Boubals D (1979) Situation des porte-greffes resistant aux nematodes ravageurs directs. Bull de L'OIV 52–578:263–271

Bouquet A, Dalmasso A (1976) Comportement d'une collection de porte-greffes de vignes en presence d'une population de nematodes (*Meloidogyne* sp.) originaire du Sud-ouest de la France. Conn Vigne Vin 10:161–174

Bourquin J-C, Otten L, Walter B (1991) Identification of grapevine rootstocks by RFLP. C R Acad Sci Paris 312:593–598

Bowers J, Boursiquot JM, This P, Chu K, Johansson H, Meredith C (1999a) Historical genetics: the parentage of Chardonnay, Gamay, and other wine grapes of northeastern France. Science 285:1562–1565

Bowers JE, Dangl GS, Meredith CP (1999b) Development and characterization of additional microsatellite DNA markers for grape. Am J Enol Vitic 50:243–246

Bredemeijer GMM, Cooke R, Ganal M, Peeters R, Isaac P, Nordijk Y, Rendell S, Jackson J, Röder MS, Wendehake K et al (2002) Construction and testing of a microsatellite database containing more than 500 tomato varieties. Theor Appl Genet 105:1019–1026

Cabezas JA, Ibanez J, Lijavetzky D, Velez D, Bravo G, Rodriguez V, Carreno I, Jermakowa AM, Carreno J, Ruiz-Garcia L et al (2011) A 48 SNP set for grapevine cultivar identification. BMC Plant Biol 11:153

Carimi F, Mercati F, Michele RD, Fiore MC, Riccardi P, Sunseri F (2011) Intra-varietal genetic diversity of the grapevine (*Vitis vinifera* L.) cultivar 'Nero d'Avola' as revealed by microsatellite markers. Genet Resour Crop Evol 58:967–975

Carvalho A, Guedes-Pinto H, Lima-Birto JE (2012) Genetic diversity in old portugese durum sheat cultivars assessed by retrotransposon-based markers. Plant Mol Biol Rep 30:578–589

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540–552

Cseh A, Taller J, Podmaniczky P, Kocsis L (2006) Comparative analysis of the most widespread grapevine rootstock lines in the world, the Teleki lines, with microsatellite markers. Cereal Res Commun 34:773–776

Doyle JJ, Doyle JA (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD et al (2008) Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452:745–749

Dzhambazova T, Hvarleva T, Hadjinicoli A, Tsvetkov I, Atanassov A, Atanassov I (2007) Characterization of grapevine rootstocks using microsatellite markers. Biotechnol Biotechnol Equip 21:58–62

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Farris JS (1989) The retention index and rescaled consistency index. Cladistics 5:417–419

Farris JS, Albert V, Källersjö M, Lipscomb D, Kluge AG (1996) Parsimony jackknifing outperforms neighbor-joining. Cladistics 12:99–124

Fatahi R, Ebadi A, Bassil N, Mehlenbacher SA, Zamani Z (2003) Characterization of Iranian grapevine cultivars using microsatellite markers. Vitis 42:185–192

Fisher BM, Salakhutdinov I, Akkurt M, Eibach R, Edwards KJ, Topfer R, Zyprian EM (2004) Quantitative trait locus analysis of fungal disease resistance factors on a molecular map of grapevine. Theor Appl Genet 108:501–515

Fossati T, Labra M, Castiglione S, Failla O, Scienza A, Sala F (2001) The use of AFLP and SSR molecular markers to decipher homonyms and synonyms in grapevine cultivars: the case of the varietal group known as 'Schiave'. Theor Appl Genet 102:200–205

Galet P (1947) Resistance des porte-greffes a la chlorose. Prog Agr Et Vit 36:128–133

Galet P (1967) Recherches sur les methods d'identification et de classification des Vitacees des zones temperees. II. These presentee a la Facalte des Sciences de Montpellier. University de Montpellier, France

Goloboff PA (1994) NONA: a tree searching program. Program and documentation. www.cladistics.com/aboutNona.htm

Grando MS, Frisinghelli C (1998) Grape microsatellite markers: sizing of DNA alleles and genotype analysis of some grapevine cultivars. Vitis 37:79–82

Granett J, Walker MA, Kocsis L, Omer AD (2001) Biology and management of grape Phylloxera. Annu Rev Entomol 46:387–412

Guetat A, Vilatersana R, Neffati M, Boussaid M (2010) Genetic diversity in Tunisian rosy garlic populations (*Allium roseum* L.) as evidenced by chloroplast DNA analysis: sequence variation of non-coding region and intergenic spacers. Biochem Syst Ecol 38:502–509

Hamilton MB (1999) Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. Mol Ecol 8:513–525

Hand ML, Cogan NOI, Forster JW (2012) Molecular characterization and interpretation of genetic diversity within globally distributed germplasm collections of tall fescue (*Festuca arundinacea* Schreb.) and meadow fescue (*F. pratensis* Huds.). Theor Appl Genet 124(6):1127–1137. doi:10.1007/s00122-011-1774-6

Harris AR (1983) Resistance of some *Vitis* rootstocks to *Xiphnema index*. J Nematol 15:405–409

Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174

Hong L, Walker MA (1998) Identifying rootstocks with simple sequence repeat (SSR) DNA markers. Am J Enol Vitic 49:403–407

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755

Hvarleva T, Rusanov K, Lefort F, Tsvetkov I, Atanassov A, Atanassov I (2004) Genotyping of Bulgarian *Vitis vinifera* L. cultivars by microsatellite analysis. Vitis 43:27–34

Ishii T, McCouch SR (2000) Microsatellites and microsynteny in the chloroplast genomes of *Oryza* and eight other Gramineae species. Theor Appl Genet 100:1257–1266

Jahnke G, Májer J, Lakatos A, Györffyné Molnár J, Deák E, Stefanovits-Bányai É, Varga P (2009) Isozyme and microsatellite analysis of *Vitis vinifera* L. varieties from the Hungarian grape germplasm. Sci Hortic 120:213–221

Jahnke G, Kocsisné Molnár G, Májer J, Szöke B, Tarczal E, Varga P, Kocsis L (2012) Analysis of grape rootstocks by SSR markers. J Int Sci Vigne Vin 45:199–210

Jansen RK, Kaittanis C, Saski C, Lee S-B, Tomkins J, Alverson AJ, Daniell H (2006) Phylogenetic analysis of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. BMC Evol Boil 6:32

Jin Y (1997) Resistance to the dagger nematode, *Xiphinema index*: screening trials and the inheritance of RAPD markers. MSc Thesis, University of California, Davis

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic, New York, pp 21–132

Khan A, Khan IA, Heinze B, Azim MK (2012) The chloroplast genome sequence of date palm (*Phoenix dactylifera* L. cv.'A-seel'). Plant Mol Biol Rep 30:666–678

Kluge AG, Farris JS (1969) Quantitative phyletics and the evolution of anurans. Syst Zool 18:1–32

Kocsis L, Lehoczky E (2000) The effect of the graperootstock-scion interaction on the Potassium and Calcium content of the leaves in connection with yield production. Commun Soil Sci Plant 31:2283–2289

Kuhsel MG, Strickland R, Palmer JD (1990) An ancient group I intron shared by eubacteria and chloroplasts. Science 250:1570–1573

Labra M, Winfield M, Ghiani A, Grassi F, Sala F, Scienza A, Failla O (2001) Genetic studies on Trebbiano and morphologically related varieties by SSR and AFLP markers. Vitis 40:187–190

Levadoux L, Boubals D, Rives M (1962) Le genre Vitis et ses especes. Ann Amelior Plant 12:19–44

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

Lowe KM, Walker MA (2006) Genetic linkage map of the interspecific grape rootstock cross Ramsey (*Vitis champinii*) × Riparia Gloire (*Vitis riparia*). Theor Appl Genet 112:1582–1592

Mandoulakani BA, Piri Y, Darishzadeh R, Bernoosi I, Jafari M (2012) Retrotransposon insertion polymorphism and genetic diversity in *Medicago sativa* populations revealed by IRAP and REMAP markers. Plant Mol Biol Rep 30:286–296

Manty F (2005) Hintergründe zur Entstehung der Bezeichnungen der Unterlagenselektionen von Sigmund Teleki und Franz Kober. Dtsch Weinbau Jahrb 57:159–164

McLeRoy SS, Renfro RE (2008) Grape man of Texas—Thomas Volney Munson—the origins of American viticulture. Wine Appreciation Guild, San Francisco, pp 76–77, 334

Moor MO (1991) Classification and systematics of eastern North American *Vitis* L. Vitaceae north of Mexico. Sida Contrib Bot 14:339–367

Moravcová K, Baránek M, Pidra M (2006) Use of SSR markers to identify grapevine cultivars registered in the Czech Republic. J Int Sci Vigne Vin 40:71–80

Müller K (2005) SeqState—primer design and sequence statistics for phylogenetic DNA data sets. Appl Bioinf 4:65–69

Nei M, Tajima F (1983) Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics 105:207–217

Nichols SA, Barnes PAG (2005) A molecular phylogeny and historical biogeography of the marine sponge genus *Placospongia* (Phylum Porifera) indicate low dispersal capabilities and widespread crypsis. J Exp Marine Biol Ecol 323:1–15

Nixon KC (2002) Winclada. Version 1.00.08. http://www.cladistics.com/aboutWinc.htm., Ithaca, NY

Nylander JAA, Wilgenbusch JC, Warren DL, Swofford DL (2008) AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24:581–583

Ogihara Y, Isono K, Kojima T, Endo A, Hanaoka M, Shiina T, Terachi T, Utsugi S, Murata M, Mori N et al (2000) Chinese spring wheat (*Triticum aestivum* L.) chloroplast genome: complete sequence and contig clones. Plant Mol Biol Rep 18:243–253

Ovesná J, Kučera L, Vaculová K, Milotová J, Snape J, Wenzl P, Huttner E, Kilian A, Martelli G, Milella L (2012) Analysis of the genetic structure of a barley collection using DNA diversity array technology (DArT). Plant Mol Biol Rep. doi:10.1007/s11105-012-0491-x

Péros J-P, Berger G, Portemont A, Boursiqout J-M, Lacombe T (2011) Genetic variation and biogeography of the disjunct *Vitis* subg. *Vitis* (Vitaceae). J Biogeogr 38:471–486

Poczai P, Taller J, Szabó I (2009) Molecular genetic study on a historical Solanum (Solanaceae) herbarium specimen collected by Paulus Kitaibel in the 18th century. Acta Bot Hung 51:337–346

Podmaniczky P, Cseh A, Taller J, Kocsis L, Győrffyné Jahnke G (2006) Genetic differences among rootstocks derived from the Teleki's seedlings. Acta Hortic 827:183–186

Pongrácz DD (1988) Rootstocks for grape-vines. Philip, Cape Town

Provan J, Russel JR, Booth A, Powell W (1999) Polymorphic chloroplast simple sequence repeat primers for systematic and population studies in the genus *Hordeum*. Mol Ecol 8:505–511

Rambaut A, Drummond AJ (2007) *Tracer* v1.4, http://tree.bio.ed.ac.uk/software/tracer/ (accessed 25 February 2012)

Ren H, Lu L-M, Soejima A, Luke Q, Zhang D-X, Chen Z-D, Wen J (2011) Phylogenetic analysis of the grape family (Vitaceae) based on the noncoding plastid *trn*C-*pet*N, *trn*H-*psb*A, and *trn*L-F sequences. Taxon 60:629–637

Renouf V, Tregoat O, Roby JP, Van Leeuwen C (2010) Soils, rootstocks and grapevine varieties in prestigious Bordeaux vineyards and their impact on yield and quality. J Int Sci Vigne Vin 44:1–8

Riaz S, Dangl GS, Edwards KJ, Meredith CP (2004) A microsatellite marker based framework linkage map of *Vitis vinifera* L. Theor Appl Genet 108:864–872

Röder MS, Wendehake K, Korzun V, Bredemeijer G, Laborie D, Bertrand L, Isaac P, Rendell S, Jackson J, Cooke RJ et al (2002) Construction and analysis of microsatellite-based database of European wheat varieties. Theor Appl Genet 106:67–73

Rossetto M, Jackes BR, Kristen DS, Henry RJ (2001) Intergeneric relationships in the Australian Vitacea: new evidence from cpDNA analysis. Genet Resour Crop Evol 48:307–314

Santibanez JM (1983) Contribution a l'etude de la resistance des Vitacees aux nematodes, Doctoral thesis, Montpellier, p 46

Saski C, Lee S, Daniell H, Wood T, Tomkins J, Kim H-G, Jansen RK (2005) Complete chloroplast genome sequence of *Glycine* max and comparative analyses with other legume genomes. Plant Mol Biol 59:309–322

Sauer MR (1967) Root-knot tolerance in some grapevine rootstocks. Austr J Agric Anim Husb 7:580–583

Schmidt J, Manty F, Cousins P (2009) Collecting *Vitis berlandieri* from native habitat sites. Acta Hortic 827:151–153

Sefc KM, Regner F, Glossl J, Steinkellner H (1998a) Genotyping of grapevine and rootstock cultivars using microsatellite markers. Vitis 37:15–20

Sefc KM, Steinkellner H, Glößl J, Kampfer S, Regner F (1998b) Reconstruction of grapevine pedigree by microsatellite analysis. Theor Appl Genet 97:227–231

Silvestro D, Michalak I (2012) raxmlGUI: a graphical front-end for RAxML. Org Divers Evol. doi:10.1007/s13127-011-0056-0

Simmons MP, Ochoterena H (2000) Gaps as characters in sequence-based phylogenetic analyses. Syst Biol 49:369–381

Simmons MP, Müller K, Norton AP (2007) The relative performance of indel-coding methods in simulations. Mol Phylogenet Evol 44:724–740

Simon D, Fewer D, Friedl T, Bhattacharya D (2003) Phylogeny and self-splicing ability of the plastid tRNA-Leu group I intron. J Mol Evol 57:710–720

Snyder E (1936) Susceptibility of grape rootstocks to rot-knot nematode. USDA Cir 405:1–16

Snyder E (1937) Grape development and improvement. USDA Yearbook of Agriculture, pp 631–664

Soejima A, Wen J (2006) Phylogenetic analysis of the grape family (Vitaceae) based on three chloroplast markers. Am J Bot 93:278–287

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analysis with thousands of taxa and mixed models. Bioinformatics 22:2688–2690

Stöver BC, Müller KF (2010) TreeGraph2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinforma 11:7

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 17:1105–1109

Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C, Willerslev E (2007) Power and limitations of the chloroplast *trn*L (UAA) intron for plant DNA barcoding. Nucleic Acids Res 35:e14

Tajima F (1989) The effect of change in population size on DNA polymorphism. Genetics 123:597–601

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:271–2739

Teleki A (1927) Der moderne Weinbau—Die rekonstruktion der Weingärten. Hertleben, Wien und Leipzig

This P, Jung A, Boccacci P, Borrego J, Botta R, Costantini L, Crespan M, Dangl GS, Eisenheld C, Ferreira-Monteiro F et al (2004) Development of a standard set of microsatellite reference alleles for identification of grape cultivars. Theor Appl Genet 109:1448–1458

Tröndle D, Schröder S, Kassemeyer H-H, Kiefer C, Koch MA, Nick P (2010) Molecular phylogeny of the genus *Vitis* (Vitaceae) based on plastid markers. Am J Bot 97:1168–1178

Upadhyay A, Saboji MD, Reddy S, Deokar K, Karibasappa GS (2007) AFLP and SSR marker analysis of grape rootstocks in Indian grape germplasm. Sci Hortic 112:176–183

Vouillamoz J, Maigre D, Meredith CP (2004) Microsatellite analysis of ancient alpine grape cultivars: pedigree reconstruction of *Vitis vinifera* L. 'Cornalin du Valais'. Theor Appl Genet 107:448–454

Wakasugi T, Sugita M, Tsuduki T, Sugiura M (1998) Updated gene map of tobacco chloroplast DNA. Plant Mol Biol Rep 16:231–241

Walker MA, Boursiquot JM (1992) Ampelographic and isozyme data correcting the misnaming of the grape rootstock S04 at the University of California, Davis. Am J Enol Vitic 43:261–265

Walker MA, Liu L (1995) The use of isozymes to identify 60 grapevine rootstocks (*Vitis* spp.). Am J Enol Vitic 46:299–305

Wang X, Ren G, Li X, Tu J, Lin Z, Zhang X (2012) Development and evaluation of intron and insertion-deletion markers for *Gossypium barbadense*. Plant Mol Biol Rep 30:605–613

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276

Weising K, Gradner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome 42:9–19

Won H, Renner SS (2005) The chloroplast *trn*T-*trn*F region in the seed plant lineage Gnetales. J Mol Evol 61:425–436

Xu H, Wilson DJ, Arulsekar S, Bakalinsky AT (1995) Sequence-specific polymerase chain-reaction markers derived from randomly amplified polymorphic DNA markers for fingerprinting grape (*Vitis*) rootstocks. J Am Soc Hortic Sci 120:714–720

Zalapa JE, Price DL, Kaeppler SM, Tobias CM, Okada M, Casler MD (2011) Hierarchical classification of switchgrass genotypes using SSR and chloroplast sequences: ecotype, ploidies, gene pools, and cultivars. Theor Appl Genet 122:805–817

Zecca G, Abbott JR, Sun W-B, Spada A, Sala F, Grassi F (2012) The timing and the mode of evolution of wild grapes (*Vitis*). Mol Phylogenet Evol 62:736–747

Zurawski G, Clegg MT (1984) The barley chloroplast DNA *atp*BE, *trn*M2, and *trn*V1 loci. Nucleic Acids Res 12:2549–2559